

Merging gravity datasets (T56)

[Top](#)

The INTREPID Gravity Merge tool can combine two vector datasets so that data points from matching groups are stored in the same group in the output dataset.

Gravity Merge is intended for combining gravity datasets where the data is grouped by station number. You can merge two datasets so that INTREPID stores the data from each station in the same group.

For general purpose merging, use the Merge Datasets tool (see [Merging datasets \(T04\)](#)).

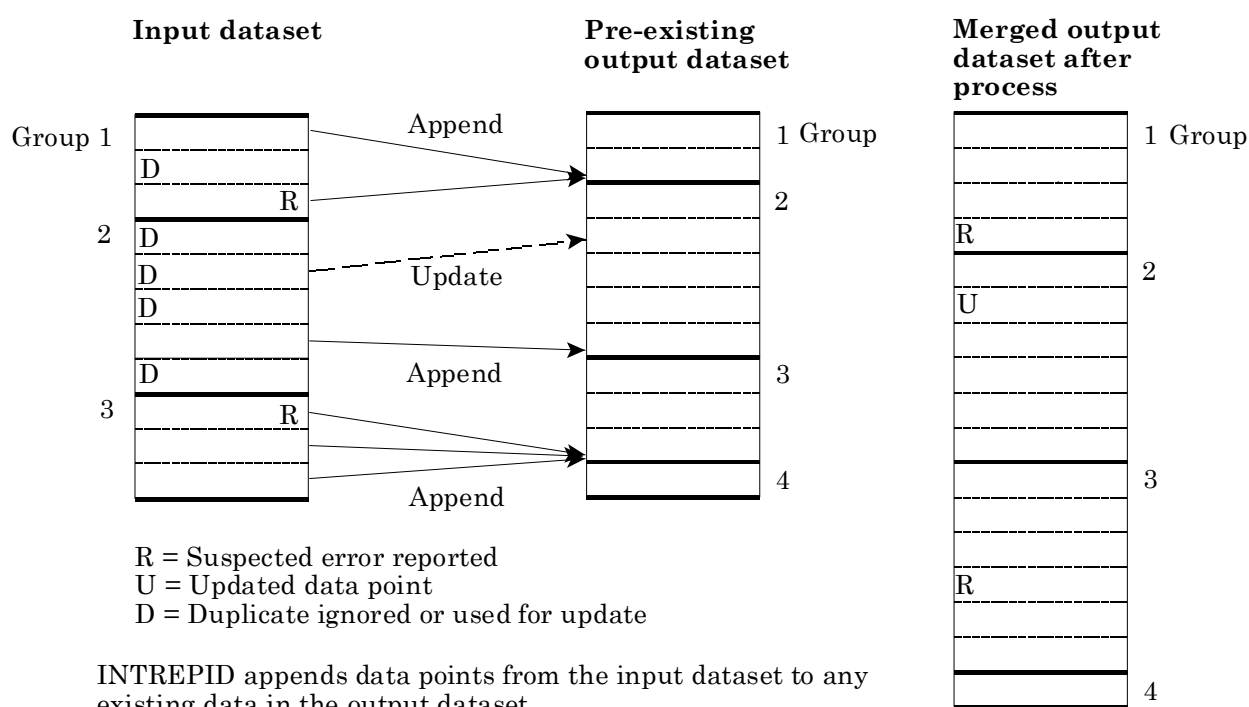
Generally speaking, both the input and output dataset must have the same fields and field names for the Gravity Merge process to work.

Gravity Merge will report:

- Eliminated duplicates
- *Nulls*
- Duplicate locations with different 'group by' (usually station number) field values
- Duplicate 'group by' (usually station number) field values with the same location field values

Here is a diagram of the merge process.

Merge process



INTREPID appends data points from the input dataset to any existing data in the output dataset.

If a point in the input dataset has fewer nulls than a duplicate point in the output dataset, INTREPID uses it for update.

INTREPID reports different locations for the same key and different keys for the same location as suspected errors.

Using the Gravity Merge tool

To use Gravity Merge with the INTREPID graphic user interface

Note: If you wish to identify duplicate locations with the same 'group by' field (usually station number), we recommend that you use geographical coordinates (latitude / longitude). These coordinates are superior for locating duplicate points spatially. See [Old Datum and Projection Conversion \(T12\)](#) for instructions on changing to this system.

- 1 Choose Merge with Master from the Gravity menu in the Project Manager, or use the command merge.exe. INTREPID displays the Gravity Merge window.
- 2 If you have previously prepared file specifications and parameter settings for Gravity Merge, load the corresponding task specification file using Load Options from the File menu. (See [Specifying input and output files](#) for detailed instructions.) If all of the specifications are correct in this file, go to step 6. If you wish to modify any settings, carry out the following steps as required.
- 3 Specify the dataset that you wish to include in the output dataset using Gravity Merge. Use Specify Input from the File menu. (See [Specifying input and output files](#) for detailed instructions.)
- 4 Specify the dataset into which you wish to insert the input dataset using Gravity Merge. Use Specify Output from the File menu. (See [Specifying input and output files](#) for detailed instructions.)
- 5 Specify the Key ('group by') field using the corresponding option from the File menu (See [Gravity Merge Key \('group by'\) field](#) for details).
- 6 When you have made specifications and settings according to your requirements, choose Apply. INTREPID will perform the Gravity Merge process and save output dataset.
- 7 If you wish to record the specifications for this process in a task specification (**.job**) file in order to repeat a similar task later or for some other reason, use Save Options from the File menu. See [Specifying input and output files](#) for detailed instructions.
- 8 If you wish to repeat the process, go to step 2.
- 9 To exit from Gravity Merge, choose Quit from the File menu.

This tool has a number of options available only if you use a task specification (**.job**) file. The headings describing these options have parentheses (for example, (Turning date stamping on or off)). A summary of the features appears in Notes in Task specification file notes and example below.

You can execute Gravity Merge as a batch task using a task specification (**.job**) file that you have previously prepared. See [Using task specification files](#) for details.

Specifying input and output files

To use Gravity Merge, you will need to specify the two vector datasets to be combined and the key field.

Choose the options as required from the File menu.

In each case INTREPID displays an Open or Save As dialog box. Use the directory and file selector to locate the file you require. (See ["Specifying input and output files" in Introduction to INTREPID \(R02\)](#) for information about specifying files).

Vector dataset notes

- INTREPID will identify X and Y fields from the dataset aliases for the process. The dataset must have the following aliases identifying appropriate fields.

Alias	Field
X	X coordinate (location)
Y	Y coordinate (location)

- If you wish to identify duplicate locations with the same 'group by' field (usually station number), we recommend that you use geographical coordinates (latitude / longitude). If you do not have a set of latitudes and longitudes available for assigning to aliases, you can create them using the Projection Conversion tool. See [INTREPID's supported datums and projections \(R09\)](#) for instructions.

See ["Vector dataset field aliases" in INTREPID database, file and data structures \(R05\)](#) for more information about aliases.

Specify Input

Use this option to specify the input dataset that you wish to include in the output dataset. This dataset should contain the second part of the data to be combined.

Specify Key Field

Use this option to specify the name of the 'group by' for linking the data in the merge process. Both datasets require the linking 'group by' field to have this name. See [Gravity Merge Key \('group by'\) field](#) for details.

Specify Output

Use this option to specify the output dataset into which you wish to insert the input dataset. This dataset may contain the first part of the data to be combined or can be a new dataset. After the process it will contain all of the data. If you specify an existing output dataset, INTREPID will build two internal hash tables to index the existing key and location fields.

Load Options

If you wish to use an existing task specification file to specify the Gravity Merge process, use this option to specify the task specification file required. INTREPID will load the file and use its contents to set all of the parameters for the Gravity Merge process. (See [Using task specification files](#) for more information).

Save Options

If you wish to save the current Gravity Merge file specifications and parameter settings as a task specification file, use this option to specify the filename and save the file. (See [Using task specification files](#) for more information).

Gravity Merge Key ('group by') field

The key ('group by') field is the field that links the datasets for the merge operation. The datasets must each have a 'group by' field with the name you specify. INTREPID combines the datasets so that data from the input dataset is included in the matching groups in the output dataset.

The most common purpose for this tool is combining gravity datasets for the same region. The normal gravity dataset 'group by' field is the station number. Gravity Merge can combine the datasets so that all of the data for a station is in the same group.

We recommend that you do not use Gravity Merge with a dataset which has the same 'group by' field value for two or more groups (e.g, if you have used the Split Group operation in the Spreadsheet Editor). Edit such a dataset using the Spreadsheet Editor before merging to ensure that all groups have unique merge key values in each dataset.

To specify the key ('group by') field:

Choose Specify Key Field from the File menu, and select the field from the input dataset.

Detecting duplicates and updating

During the merge process, INTREPID examines the X, Y and merge key (usually station number) fields and identifies suspected duplicate data points. These four fields make the default fields to compare list. If you use a task specification file you can add fields to this list using the ReportOn statement. See ([Specifying further fields to compare](#)).

(Specifying further fields to compare)

If you are using a task specification (.job) file for the Gravity Merge task you can specify additions to the fields to compare list using the **ReportOn** statement.

```
ReportOn = "fieldname, fieldname, ..."
```

Example: **ReportOn = "Elevation"**

Deleting duplicates

If the values for the fields to compare list are the same, INTREPID regards the data point as a duplicate and ignores the record in the input dataset.

Reporting suspected errors

If X and Y are the same but the merge keys are different or the merge keys are the same but X and Y are different, INTREPID retains both data points and reports them as suspected duplicates or errors.

If the difference between the X and Y is only a question of precision (number of significant figures), and two records are, in fact, duplicates, you will be able to note this from the report. You can use the Spreadsheet tool to manually delete the data point that you don't require. The number_samples() function can show you the number of data points in each group (See "[INTREPID Functions](#)" in [INTREPID expressions and functions \(R12\)](#)).

Updating with more complete data

If the values for the fields to compare list are the same but there are some nulls, INTREPID uses the data point with fewer nulls and ignores the other. (If the data point with fewer nulls is in the input dataset, then we can say that the output dataset data point is updated.)

(Merging only the last data point in each group)

You can instruct INTREPID to merge only the last data point in each group of the input dataset. This is a normal practice in the context of gravity dataset processing. This option is only available if you are using a task specification (**.job**) file.

To merge only the last record in each input dataset group

Set the **MergeLastOnly** = statement in the task specification file to YES. See [Using task specification files](#) for details.

Gravity Merge process algorithm

Here is a structured English statement of the algorithm used by the Gravity Merge tool for the merging process.

For each data point in the input dataset:

Compare it with each data point in the output dataset:

If (key field values are the same and values in the fields to compare list are the same)

Then the input data point is a duplicate

Ignore it or use it for update

Else the input data point is not a duplicate

Append it to the output dataset.

Gravity Merge reports and records

Gravity Merge reports

INTREPID prepares a report of the merge process which includes statistics on

- Duplicate data points found,
- Nulls found,
- Differences in merge key field values for the same location and
- Differences in locations for the same merge key field values.

See "[Diagnostic reporting options](#)" in [Configuring and using INTREPID \(R04\)](#) for information about process reporting in INTREPID.

Algorithm for reporting the merge process

Here is a structured English statement of the algorithm used by the Gravity Merge tool for reporting the process.

For each data point in the input dataset:

 Compare it with each data point in the output dataset:

 If (key field values are the same)

 Then

 If all fields to compare list values are the same then it is a duplicate

 Report 'data point ignored';

 If X or Y field values are not the same, then there are two different locations with the same key (usually station number).

 Report 'duplicate key'

 Else

 If X and Y field values match then there are two key values (usually station numbers) for the same location

 Report 'duplicate location';

 If (any fields to compare list values are null)

 Then

 Report 'null fields'

Date stamping merged data points

INTREPID can 'date stamp' data points as they are merged to the output dataset. If the field **HISTORYDate** exists in the output dataset, INTREPID will set it to the current date (using a "yyyy/mm/dd" string) in each data point that it appends or updates.

(Turning date stamping on or off)

If you are using a task specification file you can turn date stamping off or on. Set the **TimeStamp** = statement to **YES** or **NO** according to your requirements. See [Using task specification files](#) for details.

Apply

When you choose Apply, INTREPID will carry out the merge process that you have specified.

(Processing a specified number of data points)

If you are using a task specification (.job) file you can limit the number of records that INTREPID reads from the input dataset. This may save time when you are trialing a process.

To specify the maximum number of input dataset records to process

Use the **StopAfter** = statement in the task specification file.

Example: **StopAfter = 1000**

Exit

To exit from the Gravity Merge tool, choose Quit from the File menu.

Using task specification files

You can store sets of file specifications and parameter settings for Gravity Merge in task specification (**.job**) files.

To create a task specification file with the Gravity Merge tool

- 1 Specify all files and parameters.
- 2 If possible, execute the task (choose Apply) to ensure that it will work.
- 3 Choose Save Options from the File menu. Specify a task specification file (INTREPID will add the extension **.job**) INTREPID will create the file with the settings current at the time of the Save Options operation.

For full instructions on creating and editing task specification files see [INTREPID task specification \(.job\) files \(R06\)](#).

To use a task specification file in an interactive Gravity Merge session

Load the task specification (**.job**) file (File menu, Load Options), modify any settings as required, then choose Apply.

To use a task specification file for a batch mode Gravity Merge task

Type the command **merge.exe** with the switch **-batch** followed by the name (and path if necessary) of the task specification file.

For example, if you had a task specification file called **surv329.job** in the current directory you would use the command

```
merge.exe -batch surv329.job
```

Task specification file notes and example

Here is an example of a Gravity Merge task specification file.

Process Begin

```
Name = merge
Input = d:/survey/suppdata
Key = station_number
MergeWith = d:/survey/maindata
ReportOn = "Elevation"
MergeLastOnly = NO
TimeStamp = YES
StopAfter = 100
```

Process End

Notes

- **Input** = refers to the input dataset.

MergeWith = refers to the output dataset.

- You can specify additions to the fields to compare list using the **ReportOn** = "*fieldname, fieldname, ...*" statement. See [Detecting duplicates and updating](#) for further information.
- You can instruct INTREPID to merge only the last data point in each group of the input dataset using the **MergeLastOnly** = **YES|NO** statement. This is a normal practice in the context of gravity dataset processing. See [\(Merging only the last data point in each group\)](#) for details.
- You can turn date stamping off or on using the **TimeStamp** = **YES|NO** statement. See [Date stamping merged data points](#) for further details.
- You can limit the number of records that INTREPID reads from the input dataset using the **StopAfter** = *number_of_data_points* statement. This may save time when you are trialing a process.